AN INVESTIGATION OF SINGER PITCH DEVIATION AS A FUNCTION OF PITCH AND DYNAMICS

Perry R. Cook
Center for Computer Research in Music and Acoustics, Stanford, CA, USA

ABSTRACT

Drift and jitter were measured in singer voices, and compared across loudness and pitch, in both vibrato and non-vibrato productions. Jitter showed a slight dependence on dynamic level, and drift showed no clear dependence on dynamic level. Results correlating jitter and drift to produced pitch were more consistent if absolute sung pitch, rather than position within an individual singer's range, was used. Jitter and drift showed a slight dependence on pitch.

VOCAL PITCH DEVIATION

Deviations of pitch in the voice are important perceptual features [1][2]. Some amount of pitch deviation is present in the voice at all times, no matter how much the speaker/singer endeavors to remove it. The intentional quasi-sinusoidal modulation of the fundamental pitch is called *vibrato*, and occurs at a frequency of 5-7 Hz. in trained western BelCanto singing voices. Modulation components at frequencies higher than the vibrato are called *jitter* or *flutter*. Modulation components at frequencies lower than the vibrato rate have commonly been called *wow* or *drift*. The author prefers the terms *drift* and *jitter* because of the negative connotations of *wow* and *flutter* as distortions to be removed if possible. The production of jitter is generally regarded as an involuntary process, caused by random neural firing and a low level feedback mechanism which, in the singing voice, can be trained to cause the periodic oscillation of vibrato [3]. Drift components of very low frequency are directly related to intentional corrections in fundamental pitch. Drift is generally considered to be consciously controllable by means of an auditory feedback loop [4][3][5], but it is not possible to completely remove the drift component at will.

Most synthesis models of singer (and instrument) pitch deviation involve a single sinusoid to model the vibrato, mixed with some random signal to model both the drift and jitter components, such as simple low-pass filtered noise [6]. Maher and Beauchamp [7] proposed a more elaborate model of vocal pitch control, involving one sinusoidal oscillator, three sources of lowpass filtered noise, various summing elements, and a multiplier. The pitch perturbation research covered in this paper was conducted to investigate the behavior of the jitter and drift regions of the pitch signal spectrum as a function of sung pitch and intensity, to formulate a set of rules for pitch deviation control, and to suggest a suitable set of synthesis control parameters.

A STUDY OF SINGER JITTER AND DRIFT

Many past studies of jitter and drift have typically been conducted on tones produced by singers instructed to sing with no vibrato, because the jitter and drift components are easier to isolate and study when vibrato is absent, and many pitch detection methods yield noisy pitch estimates. Signal processing on low amplitude components in the presence of a large vibrato peak is difficult, because the jitter and drift components are often below the noise floor of the pitch detection algorithm itself [8][9]. The Periodic Predictor Pitch Tracker (PPPT) [10][11] has been shown to exhibit a noise floor of less than -55 dB relative to a sinusoidal modulation signal and -30 dB additive noise, and was used to extract the fundamental frequencies in this study. Another method [12] was used to verify the results on a randomly selected 10% of analyzed vocal tones.

Four professional singers were selected for the study, one each of the voice parts soprano, alto, tenor, and bass. The singers were instructed to sing 30 long tones on the vowel /a/ (father). Five notes were performed each at Mezzo Forte (medium loud), Pianissimo (very soft), and Fortissimo (very loud), both with and without vibrato. The singers breathed between each note, and were

allowed to repeat any notes which they felt were uncharacteristic of their ability. The frequencies produced were selected for each individual singer to evenly span that singer's comfortable range. The sound files were digitized directly to DAT, digitally transferred to computer disk, down-sampled (-96 dB stop-band rejection filter) to a sampling rate of 5512.5 Hz., then pitch signals were extracted by filtering and sampling at intervals of 55 samples. This 100 Hz. pitch signal sampling rate ensures that modulation information up to 50 Hz. was available for analysis. Once the pitch signals were obtained, Power Spectral Densities (PSDs) were calculated by performing multiple Fourier transforms in 256 point frames on each pitch signal, and averaging the magnitudes. Average and standard deviations were calculated across various groupings of spectra. To aid in generalizing characteristics of levels and rolloffs, a line was fit to the average spectra between 1 Hz. and 4 Hz, and another was fit to the region between 8 Hz. and 32 Hz.

OVERALL RESULTS

Consistent with the study of [6] was that the overall amplitude of jitter decreased with vocal range. That is, high sopranos exhibit less jitter than low basses. In the vibrato case, singers exhibited jitter spectra of about -65 dB (0.97 cents average) at 8 Hz, and rolled off at about 6 dB per octave. In the non-vibrato case, the jitter spectra were about -70 dB (0.55 cents average) at 8 Hz, and exhibited an average 8 dB per octave roll off. The standard deviations were consistently smaller in the drift region than the jitter region. The drift spectrum fell off slowly (roll off of about 1.5 dB / octave) from -50 dB (5.5 cents average) at 1 Hz. out to the vibrato peak at -50 dB average in the vibrato tones, and showed a decrease in the non-vibrato tones to -53 dB at 1 Hz. rolling off at about 2 dB per octave. This decrease implies that singers can hear their voices and control them better in the non-vibrato case than in the vibrato case, and is consistent with the model of drift as a random mechanism with control input from auditory feedback.

Dependance on Loudness

To investigate the dependence of jitter and drift on loudness, the PSD's of all pitch signals at a particular dynamic level were averaged in the vibrato and non-vibrato case. Figures 1 and 2 show plots of the PSD's of the pitch signals of all singers in the vibrato and non-vibrato cases, arranged by dynamic level. The broad dual peak nature of the aggregate vibrato peak shows the variability of vibrato rate between different singers. The average PSD jitter curves show an increase of 4 dB total from pianissimo to fortissimo. No significant change in

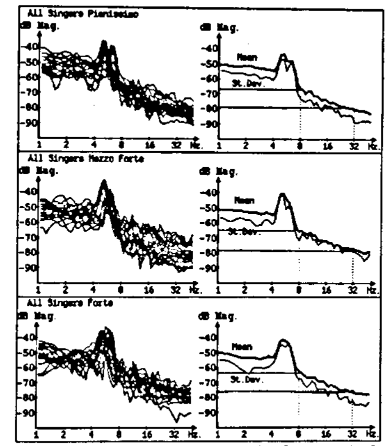

Figure 1. Power spectra of vibrato pitch signals of all singers grouped by dynamic level.

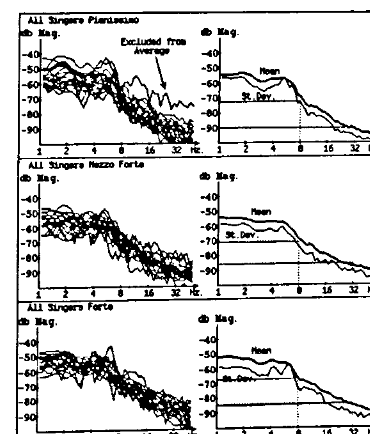

Figure 2. Power Spectra of non-vibrato pitch signals of all singers grouped by dynamic level.

spectral slope was observed, with all vibrato curves exhibiting a 6 dB/octave roll-off, and all curves without vibrato exhibiting an 8 dB/octave roll-off. The drift regions of the spectra showed no clear dependence on dynamic range, implying that the singers in this study could hear themselves and tune well at all dynamic levels.

Dependance on Pitch

To investigate how jitter and drift depend on sung pitch, two sets of spectral averages were formed. The PSD's of all singers at a particular region in their vocal range were averaged in the vibrato and non-vibrato case. Figure 3 shows the plots of the power spectral densities of the pitch signals of all singers for both vibrato and non-vibrato tones, arranged by position within the singer's range. The standard deviations of all of these plots are significantly larger than the mean spectra, indicating that grouping spectra in this way is an unreliable method of classification.

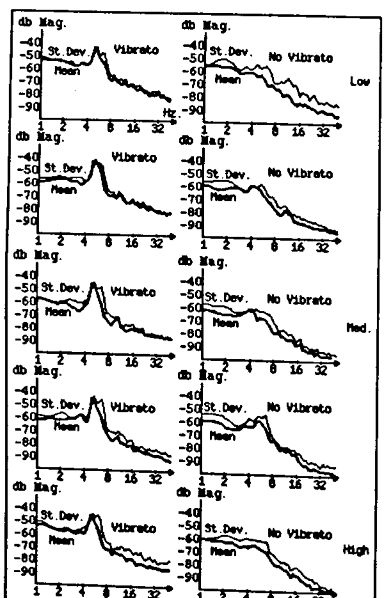

Figure 3. Singer pitch spectra averaged according to position within each singer's range.

Averages were also done within 4 one-octave frequency ranges; 90-179 Hz., 180-359 Hz., 360-719 Hz., and 720-1439 Hz. Figure 4 shows the PSD plots of the pitch signals of the singers for both vibrato and non-vibrato tones, arranged by absolute pitch. The standard deviations for these plots are quite small, indicating that the grouping of spectra by absolute pitch is a more reliable method of classification. The jitter spectra showed a slight dependence on pitch, decreasing 2 dB per octave from low pitch to high pitch. The jitter curves exhibited a consistent slope for all ranges of 8.5 dB per octave in the non-vibrato case and 6 dB per octave in the vibrato case. The drift curves showed a weak dependence on pitch, decreasing about 1 dB per octave of increasing pitch.

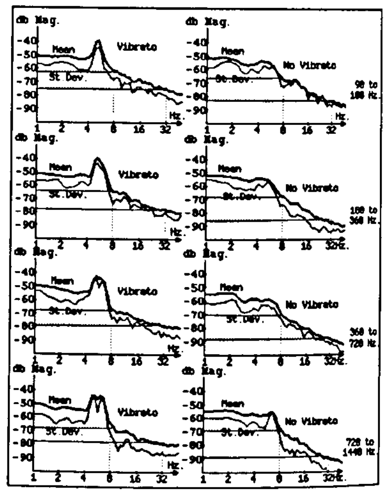

Figure 4. Singer pitch spectra averaged according to absolute pitch in octave bands.

RULES FOR SYNTHESIS

Figures 5 and 6 show the line segment approximations to the jitter and drift spectra, in the vibrato and non-vibrato cases, arranged by pitch and dynamic level. The data indicates that a suitable control space for jitter must allow control over spectral height and slope as a function of dynamic level, phonation pitch, and presence/absence of vibrato.

The minimum jitter is exhibited with no vibrato, at high pitch, and low dynamic level. This jitter is about -70 dB (.55 cents) at 8 Hz., rolling off at 8.5 dB per octave. The maximum jitter is exhibited with vibrato, at low pitch, and high dynamic level. This jitter is about -60 dB (1.7 cents), rolling off at 6 dB per octave. In both the vibrato and non vibrato case, increases in dynamic level account for about 4 dB increase in jitter across the entire dynamic range, and decreases in pitch account for about 2 dB per octave of jitter increase. From the data and the model of drift production, the drift modulation component is most strongly affected by the singer's ability to hear. An extremely simple but nearly complete model of drift is a flat spectrum at -50 dB (5.5 cents) extending to the vibrato peak. The only significant deviations from this model found in this study were in the vibrato/non-vibrato comparison, which indicated a small increase in spectral roll-off in the non-vibrato case.

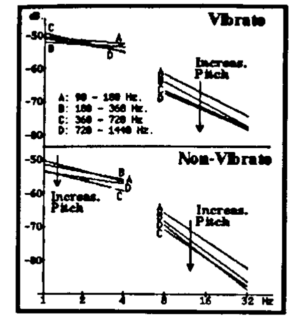

Figure 5. Line segment fits to jitter and drift spectra as function of sung pitch.

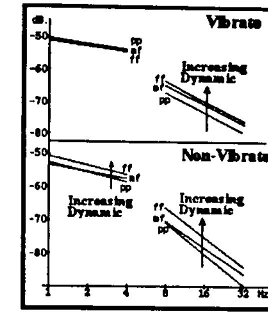

Figure 6. Line segment fits to jitter and drift as function of dynamic level.

REFERENCES

[1] Hillenbrand, J. (1970), "Perception of Aperiodicities in Synthetically Generated Voices," Journal of the Acoustical Society of America, vol. 83, no. 6, pp. 2361-2371.

[2] Kobayashi, T. & H. Sekine (1990), "Statistical Properties of Fluctuation of Pitch Intervals & its Modeling for Natural Synthetic Speech," Proc. of the IEEE Int. Conf. on Acoustics, Speech, & Signal Processing, S6a.8, pp. 321-324.

[3] Shipp, T., J. Sundberg & E. Doherty (1988), "The Effect of Delayed Auditory Feedback on Vocal Vibrato," Journal of Voice, vol. 1, no. 2, pp. 123-141.

[4] Ward, D. & E. Burns (1978), "Singing Without Auditory Feedback," Journal of Research in Singing & Applied Vocal Pedagogy, vol. 1, no. 2 pp. 24-44, 1978.

[5] Elliot, L. & A. Niemoeller (1970), "The Role of Hearing in Controlling Voice Fundamental Frequency," International Audiology, vol. 9, pp. 47-52.

[6] Ternstrom, S. & A. Friberg (1989), "Analysis and Simulation of Small Variations in the Fundamental Frequency of Sustained Vowels," Quarterly Report of the Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, no. 3, pp. 1-14.

[7] Maher, R. & J. Beauchamp (1990), "An Investigation of Vocal Vibrato for Synthesis," Applied Acoustics, vol. 30, pp. 219-245.

[8] Deem, J., W. Manning, J. Knack & J. Matesich (1989), "The Automatic Extraction of Pitch Perturbation Using Microcomputers: Some Methodological Considerations," Journal of Speech & Hearing Research, vol. 32, pp. 689-697.

[9] Hess, W. (1983), *Pitch Determination of Speech Signals*. Berlin: Springer Verlag.

[10] Cook, P. (1991), "Identification of Control Parameters in an Articulatory Vocal Tract Model, With Applications to the Synthesis of Singing," Electrical Engineering Ph.D. Dissertation., Stanford University.

[11] Cook, P., D. Morrill, & J. Smith (1993), "A MIDI Control and Performance System for Brass Instruments," Proc. of the International Computer Music Conference, pp. 130-133.

[12] Serra, X. (1989), "A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic Plus Stochastic Decomposition," Ph.D. Dissertation., Stanford University.